Hand Gestures and Posture Recognition to Control Television Using Haar-Like Features

Dewi Agushinta R. [1] Haryanto [2] Adriani Yulida K [3]
Ryan Permadi [4] Yohanes M. [5]
[1.2.3.4.5] Sarmag Program, Informatics Departement, Gunadarma University
Jl. Margonda Raya 100 Pondok Cina, Depok 16424, West Java, Indonesia
{dewiar,haryanto}@staff.gunadarma.ac.id[1, 2],
{lida_kusuma,ryan135,dewan_daru}@student.gunadarma.ac.id [3, 4, 5]

***Abstract*-Earlier time, control of television was only using a specific key or the remote. The development of ever-changing technology led to changes in a way to control the television. Changes in the way should be happen so people can interact with interactive television and user friendly. Because of the demands of time, now it no longer just a television with remote control only. But it can be also done by doing hand gesture. The hand gestures that are performed by humans will be recorded by a camera and then turned it into a command code which commands the television to do something. In this paper, we have two model structures. The first structure use computer, camera, and television. In this structure, the overall process will be performed by the computer. The second structure uses a television which has been embedded with a camera. The overall process will be performed by the television itself. In the recognition posture process, we use Haar-Like Features algorithm. This article is a proposal of idea to make a better communication between human and television.**

Keywords-computer vision; gestures database; Haar-Like features; hand gestures; posture; recognition

I. INTRODUCTION

Nowadays, all of the aspect in the world has being used the computer. It happened not only in the scope of office or school, but it has reached to daily life. Thus, we can see clearly where the interaction often happened between humans and computers. The example of this can be seen in everyday life, such as controls of the washing machine, television controls, and many more.

Because human and computer interaction was reached to entire daily life, it would require an input interface that is easy to use. There are a lot of input interface that has already exists in this world. For example the keyboard, mouse, and joystick. But sometimes it required an input interface that can read or record something in the human body. For that sort of thing, which probably will make it easier for all people include deaf people to interact with computers.

Most part of human body can be uniquely recorded for input. The concrete example is a fingerprint that is used to instruct the computer to issue permits for the users so the users can access the computer. In addition to fingerprint, there are much more parts of the human body that can be an input source. The input from the human body will be used as a code in a computer command. Human body parts include body movements, hand gestures, faces, and many more.

Although there are many parts of the human body could be used as an inputs in an input device to, but not all of these parts led to the view that is different for each acommand. This can be overcome with a hand gesture, because the hand gesture gesture can show different forms for different commands. A form of hand gesture is a command for the computer to do something.

For controlling a computer should be a real time process and place directly at the same time. Thus, the input devices is used to record the hand gesture. It also has to be able to record the hand gesture directly. Input devices that can be used directly are camera and video camera.

In recording a hand gesture, the user's hand must be recorded as a whole. If it is not recorded as a whole, it will lead to errors of the interpretation at analysis time. If the interpretation process fails, it may cause an error for computer commands.

Records of hand gesture are analyzed by an image analysis method. The results of the analysis process will be translated into an existing command codes in the database. Then the command code will be executed by a computer.

In this paper, we use a camera as an input device. This camera will record the hand gesture from the user's hand. The camera should capture the whole of user's hand. If there is a part of the user's hand does not being record, it will cause some failure in the interpretation. This failure also can lead to a command error.

The last hand gesture that is recorded as commands is intended to control the television. But it didn't just happen just that. A recorded hand gesture will be analyzed in Haar-Like Features. After getting a clear analysis of the process that is undertaken by Haar-Like Features, the next process will be translated into the

command code contained in the gesture database. When the code has been translated into commands, the code should be run by the television.

II. RELATED WORK

There is a method that explains about how to control television using hand gestures with image processing. The image processing used a normalized correlation method to trigger gesture and hand using variety of different image representations. System that they made had the ability to search multiple hand templates. The last method that they use is two tap dx and dy filters to determine the orientation of the image gradient.[1]

Real-time Vision-based Hand Gesture Recognition Using Haar-like Features [2]. This paper proposes two level approaches to solve the problem of real-time vision-based hand gesture classification. The lower level of the approach implements the posture recognition with Haar-like features and the AdaBoost learning algorithm. With this algorithm, the real-time performance and high recognition accuracy can be obtained. The higher level implements the linguistic hand gesture recognition using a context-free grammar-based syntactic analysis. Given an input gesture, based on the extracted postures, the composite gestures can be parsed and recognized with a set of primitives and production rules.

3D hand model based approaches and appearance based approaches [3]. 3 D hand model based approaches is Stenger et al presented a practical technique for model based 3D hand tracking. An anatomically accurate hand model is built from truncated quadrics. This allows for the generation of 2D profiles of the model using elegant tools from projective geometry, and for an efficient method to handle self-occlusion. The pose of the hand model is estimated with an Unscented Kalman filter (UKF). And for appearance based many method can be implemented, one example is Haar-like features, this method use image features to model the visual appearance of the hand and compare these parameters with the extracted image features from the video input.

An Extended Set of Haar-like Features for Rapid Object Detection [4]. They use haar-like features, which significantly enrich, basic set of simple haar-like features and which can also be calculated very efficiently and present a novel post optimization procedure for a given boosted cascade improving on average the false alarm rate further by 12.5%.

Vehicle Detection Method using Haar-like Feature on Real Time System [5]. They use haar-like features for Detecting vertical edge, because the vertical edge feature mostly appears at the vehicle's right and left parts. So the vertical edge provides the important information for vehicle detection. the vertical edge obtained by Haar-like feature is more robust and clearer than any other edge features (Sobel edge feature, Prewitt edge feature, etc).

III. METHOD

There are two model structures. The first model structure consist of the television combined with computer and camera. The camera used to take pictures. It will be processed by the Haar-Likes Features and become a hand posture. Then the picture is extracted and transformed into a gestures with grammar. Computer will retrieve the results of the gesture and matches it with a gestures database. Each image in the database has a command gestures respectively. For example, raised his right hand will move up one channel from the previous channel (see Fig. 1). So there are applications on the computer as interface between the hand gestures with television. However, in this paper, the application will not be discussed.

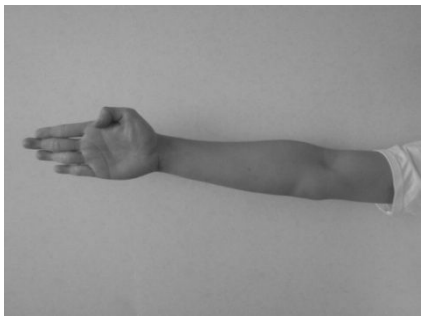

Figure 1. Example hand posture

The second model structure consists of television and camera that have been combined directly. Camera will capture images of our hands. The images that have been captured will be analyzed by the Haar-Like Features and converted into a posture. Then it will be transformed into gestures using the grammar. Gestures will be matched with the database gestures. One picture of the gestures characterize a command to the television. However, the application in this model structure already exists on the television.

Our approach is capture the image from camera. The camera will be attached in television, and process the result of Haar-Like Features algorithm. Haar-Like Features algorithm used to extract and detect image in order to get hand posture from the image, to transform posture of gesture. Since gestures are nothing but a sequence of hand postures connected by continuous motion, a recognizer can be trained against a possible grammar. With this, hand gestures can be specified as building up out of a group hand postures in various way of composition. After the gesture obtained, the gesture will be taken from many resources many people compete in order to make database gesture as possible.

Figure 2 describes the method that we use to solve the problem. First, we take the input from camera. Then the television will recognize the posture by using Haar-Like

Features. After that, the posture will be adjusted with our grammar, and then recognize the gesture. After we recognize the gesture, the program will compare the gesture with the gesture database. Then the result will be change into a command for the television using the hand gestures.

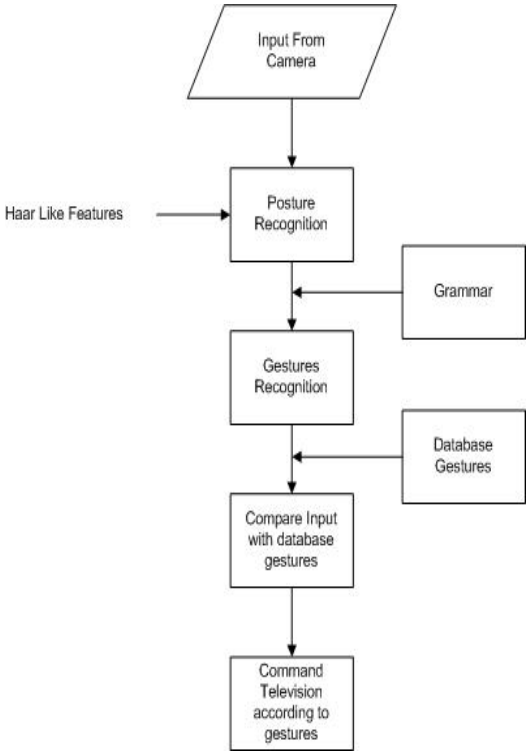

Figure 2. Our approach

To start this program, the users just need to move their hand. Then the camera will capture the movement and recognize it as s trigger. With Haar-Like features, the program will recognize the postures of users hand and then will recognize the hand gesture itself.

When the television is turning off, the camera/ video camera always stand by to capture the movements of hands. If the camera/ video camera capture the movement, the television will be turned on.

While the users move their hands, the interactive menu will appear on the television screen. The users just need move their hands like they use an invisible mouse. There is some buttons to choose the television channel or change the television volume. There is also the off buttons to turn off the television. When the users stop the hands movements, the interactive menu will be hidden. If the users want to use it, they just need to move their hands once again, and then the interactive menu will appear in the television screen again.

A. Haar-Like Features

In this paper, there are two important definitions need to be cleared. There are hand posture and hand gestures. A hand posture is defined solely by the static hand configuration and hand location without any movements involved or hand only silence and no movement. A hand gesture refers to a sequence of hand postures connected by continuous motions (global hand motion and local finger motion) over a short time span or collection of hand posture that different one of another so it will make a movement not just silence. A hand gesture is a composite action constructed by a series of hand postures that act as transition states. With this composite property of hand gestures, it is natural to decouple the problem of gesture recognition into two levels – low level posture recognition and high level gesture recognition. For hand postures, the repeatability is usually poor due to the high degree of freedom of the hand as well as the difficulty of duplicating the same working environment such as the background and the lighting condition. To solve the problem, we use a statistical approach based on a set of Haar-like features that focus more on the information within a certain area of the image rather than each single pixel [2].

Haar-Like Features is a feature of digitalize images to analyze images in object recognition. There are three kinds of methods in the Haar-Like Features algorithm, including a simple rectangular Haar-Like Features, fast Computation of Haar-Like Features, and Tilted Haar-Like Features. In this paper, we use a simple rectangular Haar-Like Features as an approach.

The simple Haar-like features (so called because they are computed similarly to the coefficients in the Haar wavelet transform) are used in the Viola and Jones algorithm. There are two motivations for the employment of the Haar-like features rather than raw pixel values. The first reason is that the Haar-like features can encode ad-hoc domain knowledge, which is difficult to describe using a finite quantity of training data. Compared with raw pixels, the Haar-like features can efficiently reduce/increase the in-class/out-of-class variability and thus making classification easier [4]. The Haar-like features describe the ratio between the dark and bright areas within a kernel. One typical example is that the eye region in the human face is darker than the cheek region, and one Haar-like feature can efficiently catch that character. The second motivation is that a Haar-like feature-based system can operate much faster than a pixel-based system. Besides the above advantages, the Haar-like features are also relatively robust to noise and lighting changes because they compute the gray level difference between the white and black rectangles. The noise and lighting variations affect the pixel values on the whole feature area, and this influence can be counteracted.

Each Haar-like feature consists of two or three connected “black” and “white” rectangles. Fig. 3 shows the extended Haar-like features set proposed by Lienhart and Maydt [4]. The value of a Haar-like feature is the difference between the sums of the pixel values within the black and white rectangles.

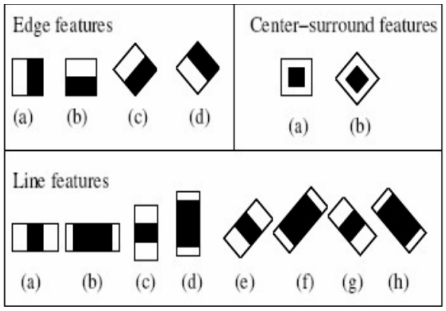

Figure 3. A set of Haar-like Features [2]

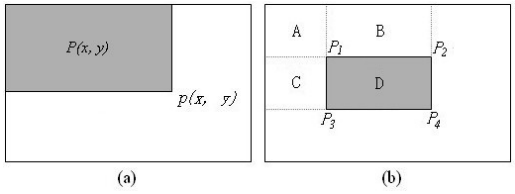

Figure 4. The Concept of Integral Image [2]

The "integral image" (Fig. 4) at location of *pixel(x, y)* contains the sum of the pixel values above and left of this pixel inclusive:

$$P(x,y) = \sum_{x' \le x, y' \le y} p(x',y') \qquad (1)$$

According to the definition of "integral image", the sum of the pixel value within the area *D* in Fig. 4(b) can be computed by:

$$P_1 + P_4 - P_2 - P_3 \qquad (2)$$

where P1 = A, P2 = A+B, P3 = A+C, and P4 = A+B+C+D as Fig. 4(b).

To detect an object of interest, the image is scanned by a sub-window containing a specific Haar-like feature. Based on each Haar-like feature *fj*, a correspondent classifier *hj(x)* is defined by:

$$h_j(x) = \begin{cases} 1, \text{if } p_j f_j(x) < p_j \theta_j \\ 0, otherwise \end{cases} \qquad (3)$$

where *x* is a sub-window, and θ is a threshold. *pj* indicates the direction of the inequality sign.

IV. CONCLUSION

Nowadays, we use remote control technology to control the television. According to the time, those technologies always develop to a better and easier technology, such as touch screen technology. But now, the world began to turn to hand gesture technology. This technology doesn't need a special monitor to control it, it just needs a camera to capture the gesture and turn it into a program to control the television.

We try to make a hand gesture control the television. We use a camera to capture the hand movements. To recognize the gesture, we use Haar Like Features algorithm. So when the user moves their hands, the camera captures the movement and then the Haar Like Feature will read the posture of the hands. After that, the posture will be recognized by using the grammar. When the program has recognized the hand gesture, it will change it into the program that is appropriate with the hand gesture that has been input before.
With this method, we can change the television channel or volume just with shaking our hands. We also can change the television configuration with move our hands.

In the future, we want try to make the prototype of this concept. We want try to add some menu to change the channel. So when the user wants to change the channel, they didn't need to move their hands according to the number. But they just have to choose the number on the screen.

REFERENCES

[1] Freeman William T., Weissman Craig D., "Television Control by Hand Gestures," Proc. IEEE Intl. Wkshp. on Automatic Face and Gesture Recognition, 1995.
[2] Chen Q., Georganas N. D., Petriu E. M., "Real-time Vision-based Hand Gesture Recognition Using Haar-like Features," Proc. IEEE Instrument and Measurement Technology Conference - 2007, 2007.
[3] Garg P., Aggarwal N., Sofat S., "Vision Based Hand Gesture Recognition," Proc. World Academy of Science, Engineering and Technology – 2009, pp. 973-977, 2009.
[4] Lienhart R., Maydt J., "An extended set of Haar-like features for rapid object detection," Proc. IEEE International Conference on Image Processing ICIP 2002, Vol. 1, pp. 900-903, 2002.
[5] Han Sungji, Han Youngjoon, Hahn Hernsoo, "Vehicle Detection Method using Haar-like Feature on Real Time System," Proc. World Academy of Science, Engineering and Technology-2009, vol. 59, pp. 455 – 459, 2009.